CHALMERS
UNIVERSITY OF TECHNOLOGY

Achievable Information Rates for Nonlinear Fiber Communication via End-to-end Autoencoder Learning

Achievable Information Rates for Nonlinear Fiber Communication via End-to-end Autoencoder Learning

Shen Li[(1)], Christian Häger[(1,2)], Nil Garcia[(1)], Henk Wymeersch[(1)]

[(1)] Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden
[(2)] Department of Electrical and Computer Engineering, Duke University, Durham, USA
shenl@student.chalmers.se

Abstract *Machine learning is used to compute achievable information rates (AIRs) for a simplified fiber channel. The approach jointly optimizes the input distribution (constellation shaping) and the auxiliary channel distribution to compute AIRs without explicit channel knowledge in an end-to-end fashion.*

Introduction

Fiber transmission rates can be increased by multi-level quadrature amplitude modulation (M-QAM) formats, which require higher input power and are thus more susceptible to nonlinear impairments such as nonlinear signal-noise interaction (NLSNI). Conventional techniques to deal with NLSNI include improved detector designs [1–3] and optimized modulation formats [3–5]. The achievable transmission rates are themselves upper-bounded by the channel capacity, which is unknown for optical channels with NLSNI, even for simplified nondispersive scenarios, though upper [6] and lower [6–8] capacity bounds have been established.

A different approach for constellation or detector design is to rely on machine learning and deep learning, including [9–15]. Recently, autoencoders (AE) have emerged as a promising tool for end-to-end design and have been shown to lead to good performance for wireless [9,10], noncoherent optical [14], as well as visible light communication [15].

In this paper, we develop an AE for a simplified memoryless fiber channel model. It is shown that the AE approach can be used to establish tight lower bounds on the channel capacity by computing achievable information rates (AIR) [16–19]. Moreover, the AE can approach maximum likelihood (ML) performance and leads to optimized constellations that are more robust against NLSNI than conventional QAM formats.

Simplified Fiber Channel Model

Similar to [1–4,6–8], we consider a simplified memoryless channel for fiber-optic communication which is obtained from the nonlinear Schrödinger equation by neglecting dispersion. The resulting per-sample model is defined by the recursion

$$x_{k+1} = x_k e^{jL\gamma|x_k|^2/K} + n_{k+1}, \quad 0 \le k < K, \quad (1)$$

where $x_0 = x$ is the (complex-valued) channel input, $y = x_K$ is the channel output, $n_{k+1} \sim \mathcal{CN}(0, P_N/K)$, L is the total link length, P_N is the noise power, and γ is the nonlinearity parameter. The model assumes ideal distributed amplification and $K \to \infty$. The channel input x is drawn randomly from an M-point constellation with $\mathbb{E}\{|X|^2\} = P_{\text{in}}$, where P_{in} is the input power.

Even though dispersive effects are ignored, the model still captures some of the nonlinear effects encountered during realistic transmission over optical fiber, in particular nonlinear phase noise (NLPN). The main interest in this channel model lies in the fact that the channel probability density function (PDF) $p(y|x)$ is known analytically [1,7,8]. This allows us to compare the AE performance to an ML detector and benchmark the obtained AIRs using known capacity bounds.

Proposed Autoencoder Structure

In machine learning, an AE is a neural network (NN) which consists of two parts: an encoder maps an input s (e.g., an image) to a lower-dimensional representation or code and a decoder attempts to reconstruct the input from the code. It has recently been proposed to interpret all components of a communication system, consisting of a transmitter, channel, and receiver, as an AE [9]. This allows for end-to-end learning of good transmitter and receiver structures.

The AE structure used in this paper is shown in Fig. 1 and will be described in the following. The goal is to transmit a message s chosen from a set of M possible messages $\{1, 2, ..., M\} \triangleq \mathcal{M}$. Following [9], the messages are first mapped to M-dimensional "one-hot" vectors where the s-th element is 1 and all other elements are 0. The one-hot vectors denoted by $\mathbf{u}$ are the inputs to a transmitter NN, which consists of multiple dense layers of neurons. Each neuron takes inputs from the

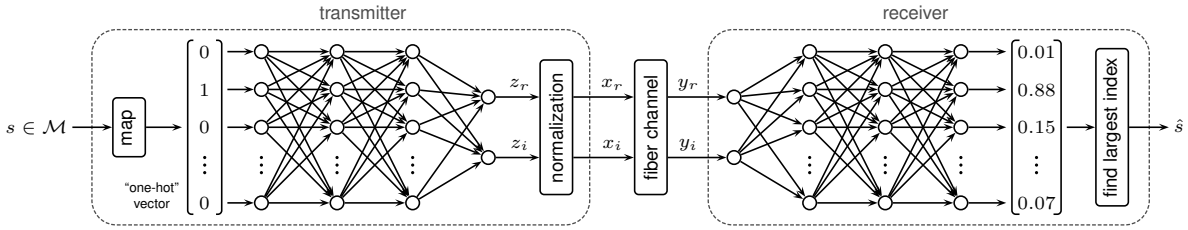

Fig. 1: Autoencoder structure assuming 2 hidden layers in both the transmitter and receiver neural network.

previous layer and generates an output according to $z_{\text{out}} = f(\mathbf{w}^{\mathsf{T}}\mathbf{z}_{\text{in}} + b)$, where $\mathbf{w}$ is a vector of weights, $b \in \mathbb{R}$ is a bias, and $f(\cdot)$ is an activation function, here considered to be a sigmoid or tanh function. The values of the two transmitter output neurons (z_r and z_i in Fig. 1) are used to form the channel input. To meet the average power constraint, a normalization is applied using M different training inputs to the NN. Then the normalized output is assumed to be sent over the channel, leading to an observation y. The real and imaginary parts of y are taken as the input to a receiver NN, the output of which we denote by $f_y(s') \in [0,1], s' \in \mathcal{M}$, where we assume a sigmoid in the last layer and then normalize the sum of the output to 1. Finally, we set $\hat{s} = \arg\max_{s'} f_y(s')$.

The AE is trained using many batches of training data averaging over different messages and channel noise configurations. In particular, the weights and biases of all neurons in both the transmitter and receiver NN are optimized with respect to $\frac{1}{N}\sum_{i=1}^{N} \ell(u_s^{(i)}, f_y(s')^{(i)})$, where

$$\ell(u_s^{(i)}, f_y(s')^{(i)}) = -u_s^{(i)} \log f_y(s')^{(i)}. \qquad (2)$$

is the cross-entropy loss, N is the batch size (a multiple of M), and the superscript refers to different training data realizations, the subscript s refers to the s^{th} element of $\mathbf{u}^{(i)}$. The optimization is performed using a variant of stochastic gradient descent with an appropriate learning rate.

Achievable Information Rates

The AE can be used to determine lower bounds on the mutual information

$$I(X,Y) = \sum_x \int p(x,y) \log_2 \frac{p(y|x)}{p(y)} \mathrm{d}y \qquad (3)$$

as follows[16–19]. We normalize $f_y(s')$ with respect to s' and consider it as a distribution over x. Then, $f_y(x)p(y)$ is a valid joint distribution over x and y, so that, due to the non-negativity of the Kullback-Leibler divergence, $\mathrm{KL}(p(x,y)||p(y)f_y(x)) \geq 0$.

Tab. 1: Autoencoder parameters

	transmitter			receiver		
layer	1	2–6	7	1	2–7	8
neurons	M	M	2	2	M	M
$f(\cdot)$	-	tanh	linear	-	tanh	sigm.

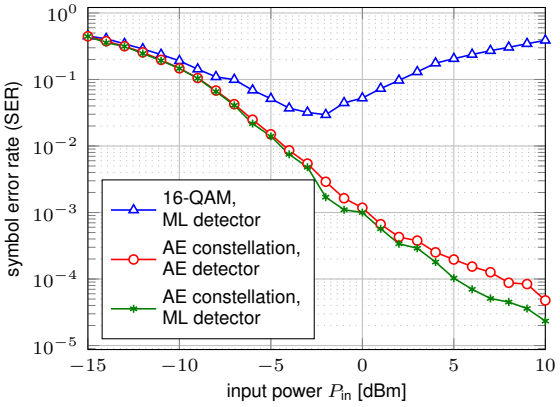

Fig. 2: SER as a function of P_{in} for $M = 16$.

Straightforward manipulations then yield

$$I(X,Y) \geq \sum_x \int p(x,y) \log \frac{f_y(x)}{p(x)} \mathrm{d}y, \qquad (4)$$

which can easily be evaluated via Monte Carlo integration. The right-hand side of (4) is the AIR of the AE. Both the mutual information and the AIR are lower bounds on the channel capacity.

Performance Analysis

For the numerical results, we assume $L = 5000$ km, $\gamma = 1.27$, and $P_N = -21.3$ dBm. The number of iterations to simulate the fiber model (1) is set to $K = 50$, which is sufficient to approximate the true asymptotic channel PDF[1]. The AE is trained separately for different values of P_{in} using the Adam optimizer in TensorFlow. The AE parameters are summarized in Tab. 1.

We start by comparing the symbol error rate (SER), i.e., $p(s \neq \hat{s})$, of the AE to the SER of an ML detector applied to (a) standard 16-QAM and (b) the signal constellation optimized by the AE. The results are shown in Fig. 2. The optimal input power for 16-QAM under ML detection is around -2 dBm, after which the SER increases due to NLPN. The SER of the AE decreases with input power, showing that the AE can find more suitable

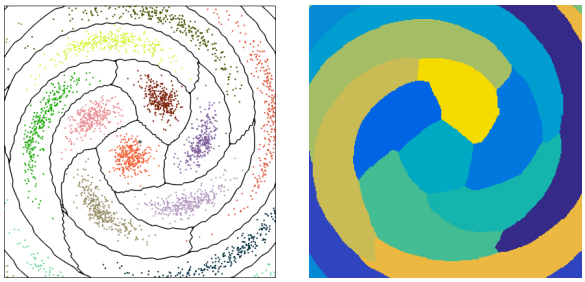

Fig. 3: ML decision boundaries for the AE constellation at $P_{\text{in}} = 0$ dBm (left) and learned AE decision regions (right).

constellations in the presence of NLPN. If we replace the receiver part of the AE with an ML detector, the SER improves only slightly. This indicates that the AE can not only learn good constellations, but also learn to approximate the correct channel distribution, thus achieving near-ML performance. To visualize this, in Fig. 3, we compare the effective decision regions implemented by the AE after training (right) to the optimal ML decision regions for the optimized AE constellation at $P_{\text{in}} = 0$ dBm (left), showing excellent agreement.

In Fig. 4, the AIR of the AE for $M = 16$ and $M = 256$ is shown. We first compare the case $M = 16$ to the mutual information $I(X;Y)$ assuming 16-QAM as the input distribution. Note that the mutual information (3) can also be evaluated via Monte Carlo integration since the channel PDF $p(y|x)$ is known. As expected, the mutual information for 16-QAM decreases with input power, whereas the AIR of the AE flattens out at the maximum value $\log_2 16 = 4$. Lastly, we compare the AIR of the AE for $M = 256$ to three information-theoretic bounds on the channel capacity: the solid black line corresponds to a recently derived upper bound[6], whereas the dashed and dash-dotted lines correspond to lower bounds based on a Gaussian[6] and half-Gaussian[8] input distribution, respectively. The AIR of the AE closely follows the maximum of the two lower bounds, slightly exceeding them at the crossover point at around 0 dBm. These results indicate that the optimized AE constellations are close to being capacity-achieving and that the upper capacity bound can be further tightened.

Conclusions

We have presented an autoencoder approach to communicating over a simplified nonlinear fiber channel. The approach allows for end-to-end learning of good signal constellations and the channel posterior distribution. It was shown that the autoencoder can learn constellations that are robust to nonlinear phase noise and outperform conventional M-QAM constellations. Moreover,

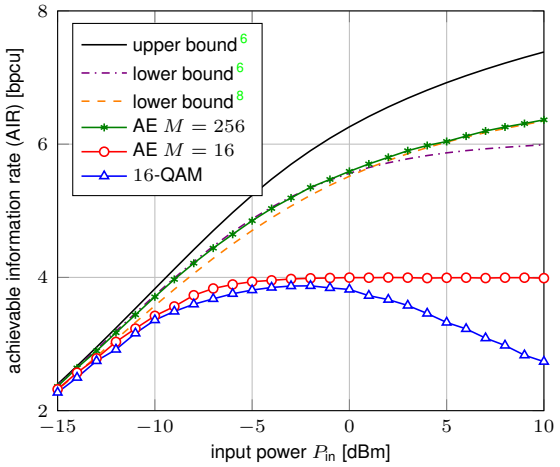

Fig. 4: Comparison of the AIR of the AE to various information-theoretic capacity bounds and 16-QAM.

near-ML performance can be obtained without explicit channel knowledge. We also evaluated the achievable information rate of the AE, showing that the obtained lower capacity bounds are comparable to, and sometimes slightly exceed, two existing lower bounds for the considered nonlinear fiber channel model.

Acknowledgements

This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 749798.

References

[1] K. P. Ho, "Phase-Modulated Optical Communication Systems", Springer (2005).
[2] A. S. Tan et al., "An ML-Based Detector for Optical Communication in the Presence of Nonlinear Phase Noise," Proc. ICC (Kyoto, Japan, 2011).
[3] A. P. T. Lau et al., "Signal Design and Detection in Presence of Nonlinear Phase Noise," J. Lightw. Technol. **25**, 3008-3016 (2007).
[4] C. Häger et al., "Design of APSK constellations for coherent optical channels with nonlinear phase noise," IEEE Trans. Commun. **61**, 3362-3373 (2013).
[5] O. Geller et al., "A Shaping Algorithm for Mitigating Inter-Channel Nonlinear Phase-Noise in Nonlinear Fiber Systems," J. Lightwave Technol. **34**, 3884-3889 (2016).
[6] K. Keykhosravi et al., "A Tighter Upper Bound on the Capacity of the Nondispersive Optical Fiber Channel," in Proc. ECOC (Gothenburg, Sweden, 2017)
[7] K. S. Turitsyn et al., "Information capacity of optical fiber channels with zero average dispersion," Phys. Rev. Lett. **91** (2003).
[8] M. I. Yousefi and F. R. Kschischang, "On the per-sample capacity of nondispersive optical fibers," IEEE Trans. Inf. Theory **57**, 7522-7541 (2011).
[9] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," in IEEE Trans. Cogn. Commun. Netw., Vol. **3**, 563-575 (2017).
[10] S. Dörner et al., "Deep Learning Based Communication Over the Air," in IEEE J. Sel. Topics Signal Process. **12**, 132-143 (2018).
[11] D. Zibar et al., "Machine learning techniques in optical communication, " J. Lightw. Techn. **34**, 1442-1452 (2016).
[12] C. Häger and H. D. Pfister, "Nonlinear Interference Mitigation via Deep Neural Networks," Proc. OFC (Los Angeles, USA, 2018).
[13] ——, "Deep learning of the nonlinear Schrödinger equation in fiber-optic communications," Proc. ISIT, (Rome, Italy, 2018).
[14] B. Karanov et al., "End-to-end Deep Learning of Optical Fiber Communications," arXiv:1804.04097 [cs.IT] (2018).
[15] H. Lee et al., "Deep learning based transceiver design for multi-colored VLC systems," Opt. Express **26**, 6222-6238 (2018).
[16] D.-M. Arnold et al., "Simulation-based computation of information rates for channels with memory," IEEE Trans. Inf. Theory **52**, 3498-3508 (2006).
[17] I. B. Djordjevic et al., "Achievable information rates for high-speed

long-haul optical transmission," J. Lightw. Techn. **23**, 3755-3763 (2005).
[18] M. Secondini et al., "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps." J. Lightw. Techn. **31**, 3839-3852 (2013).
[19] T. Fehenberger et al., "On achievable rates for long-haul fiber-optic communications," Opt. Express **23**, 9183-9191 (2015).